Original Research Article

Machine Learning-Based Predictive Maintenance for Manufacturing: Optimizing Failure Detection with Sensor-Based Models

Aadhav Hariish

Heritage High School, 14040 Eldorado Pkwy, Frisco, TX 75035, United States

ABSTRACT

Reducing unplanned equipment failures is critical for improving efficiency and lowering operational costs in manufacturing systems. This study investigates the use of machine learning models to predict machine failure using sensor-based operational data. The AI4I 2020 Predictive Maintenance Dataset was used, which includes temperature, rotational speed, torque, tool wear, and machine type variables. Exploratory data analysis identified key patterns, including nonlinear relationships between torque and failure probability and strong correlations among certain sensor variables. Four machine learning models, Logistic Regression, Random Forest, Gradient Boosting, and Support Vector Machine (SVM), were developed and evaluated. Due to class imbalance, performance was assessed using precision, recall, and F1-score, with emphasis on failure detection. While most models achieved high accuracy (0.97-0.99), Logistic Regression performed poorly in identifying failures (recall = 0.12). SVM achieved a high recall (0.84) but suffered from low precision (0.27), resulting in excessive false positives. A tuned Gradient Boosting model with threshold optimization achieved the best overall performance, with a precision of 0.77, a recall of 0.74, and an F1-score of 0.75. Receiver operating characteristic (ROC) analysis showed strong model discrimination, with an area under the curve (AUC) of 0.96 for Gradient Boosting and 0.95 for Random Forest. Compared to the untuned model, recall improved substantially, enabling the detection of more failure events with a moderate increase in false positives. These results demonstrate that ensemble methods, combined with threshold tuning, provide an effective approach for predictive maintenance by balancing failure detection and false alarm rates.

Keywords: Machine Learning; Predictive Maintenance; Failure Prediction; Sensor Data Analysis; Gradient Boosting; Industrial Data Analytics

INTRODUCTION

Manufacturing systems are under increasing pressure to improve productivity while managing rising labor, material, maintenance, and downtime-related costs. Recent industry reports indicate that manufacturers continue to face higher operating expenses, supply chain uncertainty, and operational disruptions that reduce profitability and efficiency (1, 2). Unplanned equipment failures are particularly costly because they interrupt production, increase maintenance expenses, and can negatively affect product quality. As a result, manufacturers are seeking more effective methods to improve reliability while controlling costs.

Traditional maintenance strategies rely heavily on scheduled inspections and time-based servicing.

Corresponding author: Aadhav Hariish, E-mail: aadhav0404@gmail.com.

Accepted June 6, 2026
https://doi.org/10.70251/HYJR2348.43410418

Although preventive maintenance can reduce unexpected breakdowns, it often results in unnecessary maintenance activities and may fail to identify equipment degradation between scheduled inspections. Consequently, organizations are increasingly adopting data-driven approaches that improve maintenance efficiency, reduce operational downtime, and enhance productivity (3-5).

Machine learning offers a promising solution through predictive maintenance, where models analyze sensor data such as temperature, rotational speed, torque, and tool wear to identify conditions associated with future failures. By enabling maintenance decisions based on actual equipment condition rather than fixed service schedules, predictive maintenance can reduce downtime and maintenance costs. Previous research has demonstrated the effectiveness of machine learning for failure detection. For example, Hosseinzadeh et al, (2023) used a Gradient Boosted Decision Tree model to analyze predictive maintenance data and achieved an accuracy and recall of 0.93, demonstrating the potential of machine learning to identify early failure conditions before major equipment breakdowns occur (6).

This study utilizes the AI4I 2020 Predictive Maintenance Dataset, which contains 10,000 observations of machine operation and includes sensor measurements such as air temperature, process temperature, rotational speed, torque, and tool wear, along with machine failure outcomes (7). The dataset provides a realistic representation of industrial operating conditions and enables the evaluation of machine learning models using both sensor measurements and labeled failure events.

The primary research question of this study is: How can machine learning models improve machine failure detection for predictive maintenance? A secondary research question investigated which sensor variables provide the most actionable insight for identifying failure risk. To address these questions, four machine learning approaches, Logistic Regression, Random Forest, Gradient Boosting, and Support Vector Machine, were developed and evaluated. Model tuning and threshold optimization were subsequently applied to identify the approach that provides the most effective balance between failure detection and false alarm rates.

METHODS AND MATERIALS

Dataset Description

This study utilizes the AI4I 2020 Predictive Maintenance Dataset, obtained from the UCI Machine Learning Repository (7). The dataset contains 10,000 observations of machine operation, with each observation representing a single instance of equipment performance under varying operating conditions.

The dataset includes both continuous sensor distributions and categorical variables relevant to machine operation. The primary input variables consist of air temperature, process temperature, rotational speed, torque, and tool wear, which collectively describe the mechanical and environmental state of the machine. In addition, a categorical variable, machine type (L, M, H), represents different machine quality levels - low, medium, and high - which provides additional context on machine configuration and contributes to variation in failure risk.

The target variable, machine failure, is binary and indicates whether a failure event occurred. Additional failure subtype variables (e.g., tool wear failure, heat dissipation failure) are included in the dataset but were excluded from modeling to prevent data leakage, as they are derived from the primary failure outcome.

This dataset is well-suited for predictive maintenance applications, as it combines real-time sensor data with machine-specific characteristics, enabling the development of machine learning models to identify patterns associated with increased failure risk.

Data Visualization and Exploratory Analysis

Before developing any models from the AI4I 2020 Predictive Maintenance Dataset, exploratory data analysis (EDA) was used to understand the dataset better and find patterns that might indicate when a machine could experience failure. Histograms of all sensor variables were created to help identify the distributions of the main sensor variables. By examining the failure rate of machines by type, categorical differences in failure risk could be assessed. Visualizations on torque were created to analyze the relationship between torque and the probability of machine failure. To better understand the relationships among the main sensor variables, a pairwise plot was created, and all visualizations were produced using EDA to inform feature selection and the selection of machine learning models capable of modeling nonlinear relationships.

Data Preprocessing

The dataset was first cleaned by removing non-predictive and leakage-prone variables, including failure subtype indicators. The target variable was defined as machine failure, and the remaining variables were used as input features.

Categorical variables, specifically machine type, were converted into numerical form using one-hot encoding. Continuous sensor variables were standardized using feature scaling to ensure consistent input for models sensitive to feature magnitude.

The dataset was then split into training and testing sets using an 80/20 stratified split to preserve the class imbalance between failure and non-failure observations.

Model Development

Four machine learning models were developed and evaluated: Logistic Regression, Random Forest, Gradient Boosting, and Support Vector Machine (SVM). These models were selected to represent a range of approaches, including linear, tree-based, and kernel-based methods. Model performance was evaluated using accuracy, precision, recall, and F1-score, with particular emphasis on recall and F1-score for the failure class due to class imbalance and the importance of detecting failure events.

Model Optimization

Hyperparameter tuning was performed using grid search with cross-validation. For tree-based models, parameters such as the number of estimators, learning rate, and tree depth were optimized. Model selection prioritized recall to improve the detection of rare failure events.

Following model tuning, threshold optimization was applied to adjust the classification cutoff. Instead of using the default threshold of 0.5, multiple thresholds were evaluated to identify the optimal balance between precision and recall. The final model, a tuned Gradient Boosting classifier, used a threshold of 0.2, which maximized the F1-score while significantly improving recall.

Model Evaluation

The final model was evaluated using confusion matrices, receiver operating characteristic (ROC) curves, and precision-recall curves. These evaluation methods provided insight into model performance across different thresholds and ensured that the selected model effectively balanced failure detection with minimizing false positives.

RESULTS AND DISCUSSION

Exploratory Data Analysis Results

The distributions of the key sensor variables are shown in Figure 1. Air temperature and process

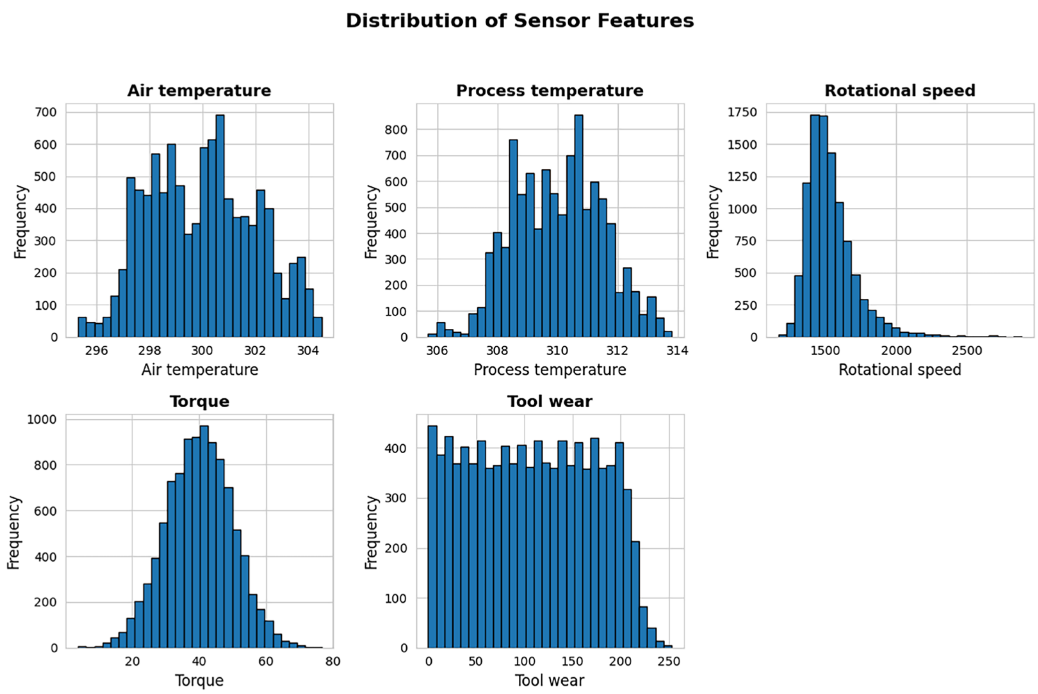

Figure 1. Distributions of key sensor variables, including air temperature, process temperature, rotational speed, torque, and tool wear, across all machine operations. These distributions characterize typical operating conditions and provide a baseline for identifying deviations that may be associated with machine failure events.

temperature exhibit approximately normal distributions, indicating stable ambient and operating conditions. Torque follows a similar pattern, suggesting that most observations fall within a typical operating load range. In contrast, rotational speed is somewhat skewed toward lower values, while tool wear displays a more uniform distribution, reflecting gradual accumulation over time. Together, these distributions establish a baseline for normal machine behavior, where deviations from these patterns may indicate abnormal operating conditions or potential degradation.

The relationship between machine type and failure rate is illustrated in Figure 2. Type L machines exhibit the highest failure rate (approximately 3.9%), followed by Type M (approximately 2.8%) and Type H (approximately 2.1%). This demonstrates that machine configuration contributes to variation in failure risk and highlights the importance of incorporating categorical variables alongside sensor data in predictive models.

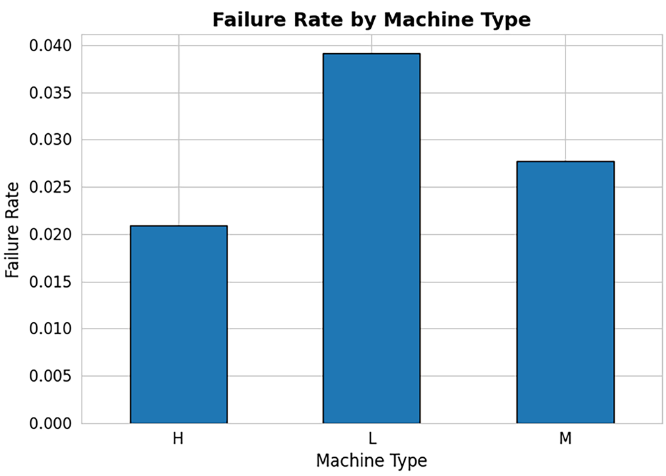

Figure 2. Failure rate by machine type low (L), medium (M), high (H), illustrating variation in failure risk across machine configurations. This result highlights the value of incorporating categorical features, such as machine type, alongside sensor data in predictive models.

The relationship between torque and failure probability is presented in Figure 3 (A). Failure probability remains low at lower torque levels but increases sharply beyond a mid-range threshold, indicating a nonlinear relationship between torque and machine failure. This suggests that higher torque levels correspond to increased mechanical stress and a greater likelihood of failure, making torque a critical predictor in the modeling process.

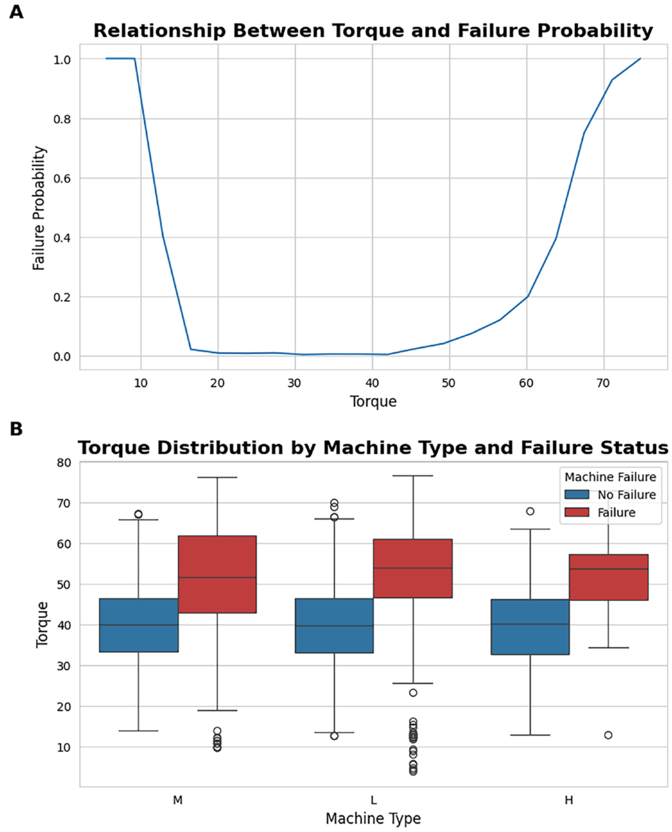

Figure 3. (A) Relationship between torque and failure probability, showing a nonlinear increase in failure risk at higher torque levels. (B) Distribution of torque by machine type (L, M, H) and failure status, showing higher median torque values for failed machines across all machine types. Together, these results highlight torque as a key predictor of machine failure.

The pairwise relationships among the sensor variables are shown in Figure 4.

A strong positive linear relationship is observed between air temperature and process temperature, indicating that these variables are highly correlated and may provide redundant information in predictive models.

In contrast, torque and rotational speed exhibit a clear inverse nonlinear relationship. As rotational speed increases, torque decreases in a curved pattern, reflecting the mechanical trade-off between these operational variables. Notably, regions characterized by lower rotational speed and higher torque show a greater concentration of failure events, suggesting that these operating conditions are associated with increased failure risk.

From a modeling perspective, these relationships

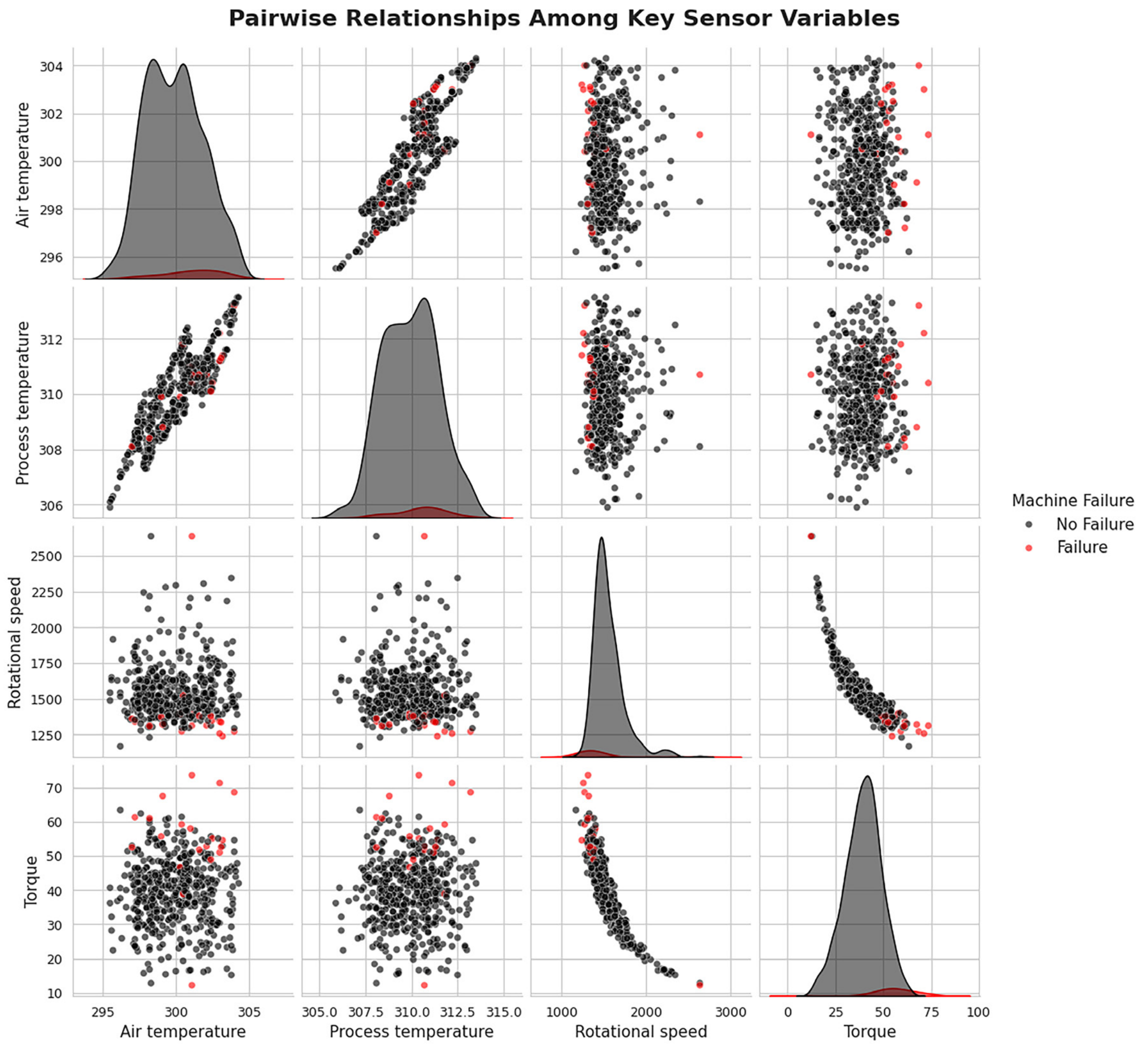

Figure 4. *Pairwise relationships among sensor variables (air temperature, process temperature, rotational speed, torque, and tool wear), with observations colored by failure status. The matrix highlights correlations and nonlinear interactions that inform feature selection and model choice.*

highlight the importance of feature selection and model choice. The strong correlation between air and process temperatures suggests that one of these variables may be redundant, while the nonlinear interaction between torque and rotational speed indicates the need for models capable of capturing complex relationships. Tree-based models, such as Random Forest and Gradient Boosting, are well-suited for this purpose and can improve predictive performance by modeling these interactions.

These visual analyses highlight nonlinear relationships and differences across machine types, supporting the use of machine learning models capable of capturing complex feature interactions.

Model Performance Results

The performance of the machine learning models developed in this study was evaluated using accuracy, precision, recall, and F1-score, with emphasis on failure detection due to class imbalance. A comparison of the results of all the models is presented in Table 1.

Table 1 shows the results from several machine learning algorithms used to predict machine failure. Although most models achieved high accuracy scores (0.97-0.99), accuracy alone is not sufficient due to the highly imbalanced dataset, where failure instances are rare compared to non-failure instances.

The Logistic Regression model achieved high

Table 1. *Performance comparison of machine learning models for machine failure prediction, evaluated using accuracy, precision, recall, and F1-score.*

Model	Accuracy (Failure)	Precision (Failure)	Recall (Failure)	F1-score (Failure)
Logistic Regression	0.97	0.67	0.12	0.20
Random Forest	0.98	0.89	0.37	0.52
Gradient Boosting	0.98	0.89	0.49	0.63
SVM	0.99	0.27	0.84	0.41
Tuned Gradient Boosting	0.98	0.77	0.74	0.75

accuracy (0.97) but performed poorly in identifying failures, with a recall of 0.12 and an F1-score of 0.20. This indicates that, while the model correctly classifies most non-failure cases, it fails to detect failure events effectively. In contrast, ensemble models such as Random Forest and Gradient Boosting were able to capture nonlinear relationships in the data, resulting in improved recall and F1-scores.

The SVM achieved the highest recall (0.84), demonstrating a strong ability to detect failures. However, this came at the cost of very low precision (0.27), indicating a large number of false positives. Such behavior limits its practical applicability in predictive maintenance, where excessive false alarms can lead to unnecessary maintenance actions.

The tuned Gradient Boosting model achieved the best overall performance, with a precision, recall, and F1-score of 0.77, 0.74, and 0.75, respectively. Compared to the untuned model, recall improved significantly (from 0.49 to 0.74), while precision decreased moderately (from 0.89 to 0.77). This reflects a deliberate tradeoff, where the model prioritizes detecting more failures at the expense of a modest increase in false positives. Importantly, this tradeoff is favorable in predictive maintenance, where missed failures are more costly than unnecessary inspections.

The performance of the optimized Gradient Boosting model is further illustrated in Figure 5, which presents the confusion matrix at a threshold of 0.2.

Figure 5 shows that the model correctly identifies 50 failure cases while missing 18 failures. This represents a substantial improvement over earlier models, which missed a significantly larger number of failure events. However, the improved recall comes with an increase in false positives, with 15 non-failure cases incorrectly classified as failures. This highlights the inherent tradeoff between sensitivity and precision. Increasing recall is critical, since undetected failures can result in significant operational downtime and cost, whereas false positives typically incur lower costs.

The relationship between precision and recall across different thresholds is shown in Figure 6, which compares the tuned Gradient Boosting and Random Forest models.

The precision-recall curves demonstrate that the Gradient Boosting model maintains higher precision across a wide range of recall values compared to the Random Forest model. As recall increases, both models exhibit a decline in precision due to the inherent tradeoff between these metrics. However, the decline is less pronounced for the Gradient Boosting model, indicating that it is better able to detect failures without

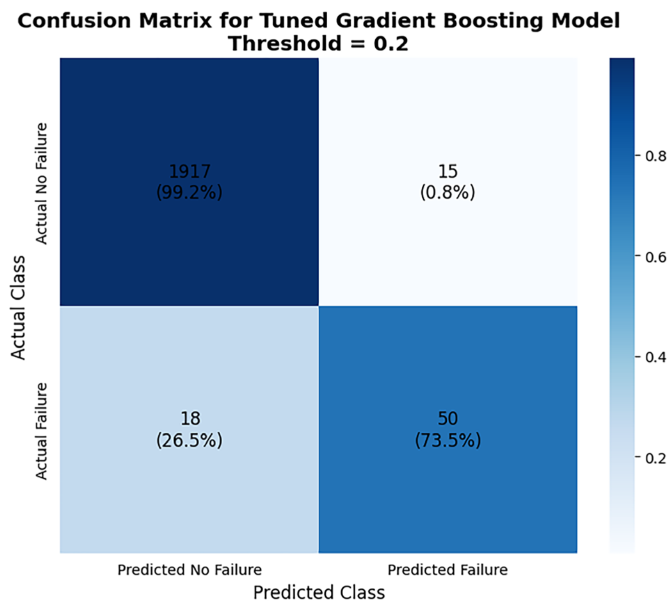

Figure 5. *Confusion matrix for the tuned Gradient Boosting model at a threshold of 0.2, showing improved detection of failure cases with 50 correctly identified failures and 18 missed failures.*

being overwhelmed by false positives. This reinforces the conclusion that Gradient Boosting provides a more effective balance between detection and reliability.

Finally, the overall classification performance across all thresholds is shown in Figure 7, which presents the ROC curves for both models.

Both models demonstrate strong performance, with AUC values of 0.96 for Gradient Boosting and 0.95 for Random Forest. However, the Gradient Boosting curve consistently lies above that of Random Forest, indicating superior performance across all classification thresholds. This means that, for any given false positive rate, the Gradient Boosting model achieves a higher true positive rate than the Random Forest model, further confirming its effectiveness in identifying machine failures.

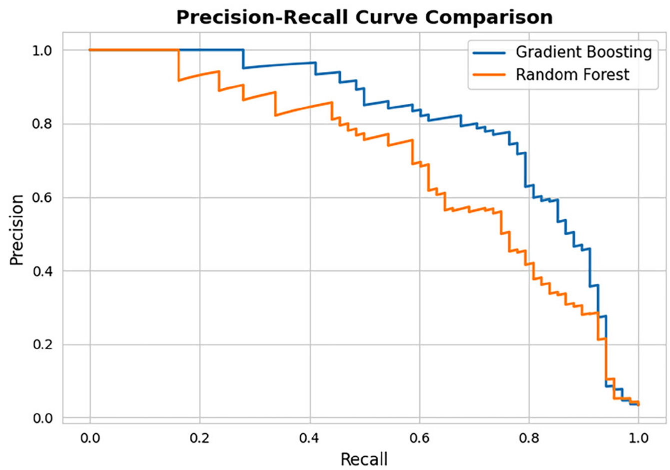

Figure 6. *Precision–recall curves for the tuned Gradient Boosting and Random Forest models, illustrating the tradeoff between precision and recall across classification thresholds.*

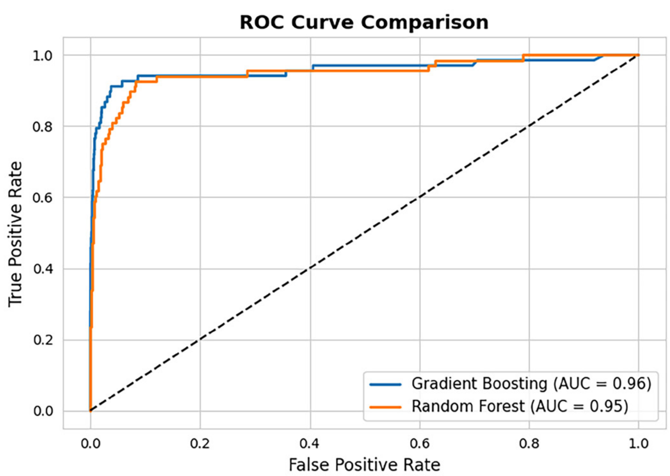

Figure 7. *Receiver operating characteristic (ROC) curves for the tuned Gradient Boosting and Random Forest models, with corresponding area under the curve (AUC) values.*

Overall, the results demonstrate that the tuned Gradient Boosting model provides the best balance between failure detection and false alarm rates. Its superior performance across multiple evaluation metrics and thresholds makes it well-suited for predictive maintenance applications, where accurate and timely identification of potential failures is critical.

The relative importance of each input variable in the tuned Gradient Boosting model is shown in Figure 8. Feature importance analysis provides a quantitative measure of how much each variable contributes to machine failure prediction and helps address the secondary research question concerning which sensor measurements provide the most actionable information for predictive maintenance.

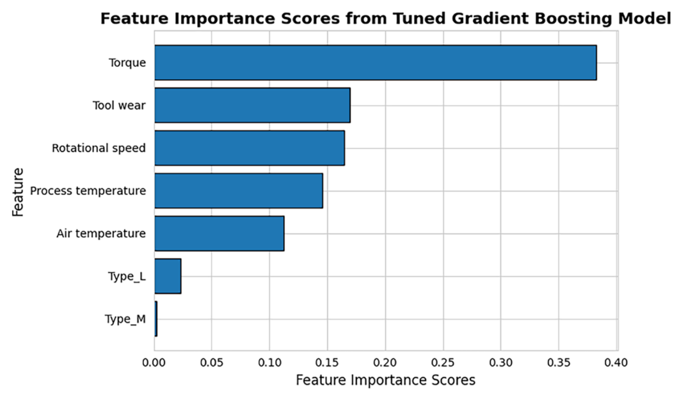

Figure 8. *Feature importance scores from the tuned Gradient Boosting model, showing the relative contribution of each input variable to machine failure prediction. Higher importance values indicate a greater influence on model predictions.*

The results indicate that torque and rotational speed are among the most influential predictors of machine failure, supporting the patterns observed during exploratory data analysis. Earlier visualizations showed that failure probability increased substantially at higher torque levels and that torque exhibited a nonlinear relationship with rotational speed. The feature importance analysis confirms that these variables contribute strongly to model performance and are key indicators of machine operating conditions associated with elevated failure risk.

In contrast, air temperature and process temperature contribute less independently to prediction accuracy, likely because these variables are highly correlated and therefore provide overlapping information. Overall, the feature importance results provide quantitative evidence that operational load-related variables, particularly torque and rotational speed, play a central role in machine failure prediction and should be prioritized when developing predictive maintenance strategies.

CONCLUSION

This study examines the use of machine learning models to predict machine failure and support more efficient maintenance strategies. Using the AI4I 2020 Predictive Maintenance Dataset, multiple models were developed and evaluated, including Logistic Regression, Random Forest, Gradient Boosting, SVM, and a tuned Gradient Boosting model.

The results demonstrate that the Gradient Boosting model achieved the best overall performance by providing a strong balance between precision and recall. Through threshold optimization (0.2), the model significantly improved its ability to detect failure events, increasing recall while maintaining acceptable precision. This tradeoff is particularly valuable in predictive maintenance, where missing a failure can result in substantial operational costs.

Exploratory analysis and feature importance evaluation identified torque and rotational speed as the most influential predictive features, with their nonlinear relationship contributing significantly to failure risk. These findings highlight the importance of incorporating both feature interactions and nonlinear modeling techniques when developing predictive maintenance systems.

Despite these promising results, several limitations should be considered. First, the dataset represents simulated or controlled operating conditions and may not fully capture the complexity of real-world manufacturing environments. Second, failure events are relatively rare, which introduces class imbalance and may affect model generalization. Future work could incorporate real-time industrial data, additional contextual variables, and more advanced modeling techniques to further improve predictive accuracy and practical applicability.

Overall, this study demonstrates that machine learning, particularly tuned ensemble methods, can effectively support predictive maintenance by improving failure detection while balancing false alarm rates. Improved failure detection may help manufacturers make more informed maintenance decisions and reduce the operational impact of unexpected equipment failures.

ACKNOWLEDGMENT

The author would like to express sincere thanks to Dr. John Basbagill, who is a Postdoctoral Researcher at Stanford University, for his guidance and mentorship during the course of this project. His ability to connect theoretical machine learning principles to practical applications was especially valuable in shaping this project. His mentorship played a vital role in developing an enhanced perspective towards implementing predictive models in manufacturing processes.

CONFLICT OF INTEREST

The author declares no conflicts of interest related to this work.

REFERENCES

1. Deloitte. 2025 Manufacturing Industry Outlook. Available from: https://www.deloitte.com/us/en/insights/industry/manufacturing-industrial-products/manufacturing-industry-outlook.html (accessed on 2026-04-13).
2. World Manufacturing Foundation. The 2024 World Manufacturing Report: New Perspectives for the Future of Manufacturing Outlook 2030. Available from: https://worldmanufacturing.org/wp-content/uploads/14/6-WM-REPORT_2024_LD_E-Book_b_Final.pdf (accessed on 2026-11-23).
3. McKinsey & Company. Maintenance and operations: Is asset productivity broken? Available from: https://www.mckinsey.com/industries/electric-power-and-natural-gas/our-insights/maintenance-and-operations-is-asset-productivity-broken#/ (accessed on 2026-04-02).
4. Gregolinska E, Khanam R, Lefort F, Parthasarathy P. Capturing the true value of Industry 4.0. Available from: https://www.mckinsey.com/capabilities/operations/our-insights/capturing-the-true-value-of-industry-four-point-zero (accessed on 2026-03-29).
5. Küpper D, Knizek C, Ryeson D, Nöcker J. Quality 4.0 Takes More Than Technology. Available from: https://

www.bcg.com/publications/2019/quality-4.0-takes-more-than-technology (accessed on 2026-04-13).

6. Hosseinzadeh A, Chen FF, Shahin M, Bouzary H. A predictive maintenance approach in manufacturing systems via AI-based early failure detection. *Manufacturing Letters*. 2023 Aug 1; 35: 1179-1186. Available from: https://www.sciencedirect.com/science/article/pii/S2213846323001864 (accessed on 2026-04-10). https://doi.org/10.1016/j.mfglet.2023.08.125
7. AI4I 2020 Predictive Maintenance Dataset. UCI Machine Learning Repository. Available from: https://doi.org/10.24432/C5HS5C (accessed on 2026-04-10).